سومین کنفرانس فناوری اطلاعات و دانش
۶ تا ۸ آذرماه ۱۳۸۶

دسته بندی داده های مربوط به گوجه فرنگی و بررسی و ارزیابی این روشها

مرتضی صابری کمرپشتی
دانشگاه آزاد اسلامی فیروزکوه
Msaberyk@yahoo.com

حمیدرضا پوررضا
دانشگاه فردوسی مشهد
Hrpourreza@um.ac.ir

چکیده – هدف ما در این مقاله دسته بندی داده های مربوط به گوجه فرنگی است . داده های مورد استفاده در این مقاله از ۹۰ عکس از محصولات گوجه فرنگی می باشد که از این تعداد عکس ما با استفاده از تکنیکهای بینایی ماشین حدود ۱۰ خصوصیت را استخراج کردیم. در این مقاله به بررسی و ارزیابی خصوصیات بدست آورده شده از این عکس ها می پردازیم و همچنین این داده ها را با استفاده از انواع مختلفی از دسته بندی کننده ها دسته بندی می کنیم و نتایج حاصل را با هم مقایسه می کنیم .

کلید واژه- دسته بندی , گوجه فرنگی , شبکه عصبی , پردازش تصویر .

۱- مقدمه

در صنایع کشاورزی از کاربردهای پردازش تصویر استفاده های زیادی شده است که می توان به استفاده از آن برای دسته بندی کیفی سیب , سیب زمینی , فلفل , خیار , گوجه وغیره اشاره کرد[6][5][4] .ما در این مقاله بر روی دسته بندی داده های مربوط به گوجه فرنگی متمرکز شده ایم . در این زمینه کارهای متفاوتی انجام شده است . برخی از این مقالات بر روی دسته بندی گوجه فرنگی بر اساس رنگ آن متمرکز شده اند بدین صورت که بر اساس مقدار سبزی یا قرمزی گوجه ها آنها را دسته بندی می کنند [1][2][3].

برخی دیگر از این مقالات نیز دسته بندی کیفی این داده ها را با استفاده از تکنیکهای شبکه عصبی و منطق فازی انجام داده اند[7]. در این مقاله سعی کرده ایم که عمل دسته بندی داده های مربوط به گوجه فرنگی را با استفاده از دسته بندی کننده های مختلف انجام دهیم و نتایج آن را با هم مقایسه کنیم .

۲- شرح تکنیکی

ما در این مقاله قصد داریم تا این گوجه فرنگی ها را در سه دسته خوب , متوسط و بد دسته بندی کنیم. داده های مورد استفاده در این مقاله از ۹۰ عکس از محصولات گوجه فرنگی جمع آوری شده است که از این تعداد عکس ما با استفاده از تکنیکهای بینایی ماشین تعداد ۱۰ خصوصیت را استخراج کردیم . این خصوصیات عبارتند از :

۱. سبز بودن
۲. قرمز بودن
۳. زرد بودن
۴. میانگین زردی و قرمزی و سبزی
۵. آنتروپی
۶. انرژی
۷. کنتراست
۸. همگنی

۹. گردی

۱۰. مساحت گوجه فرنگی

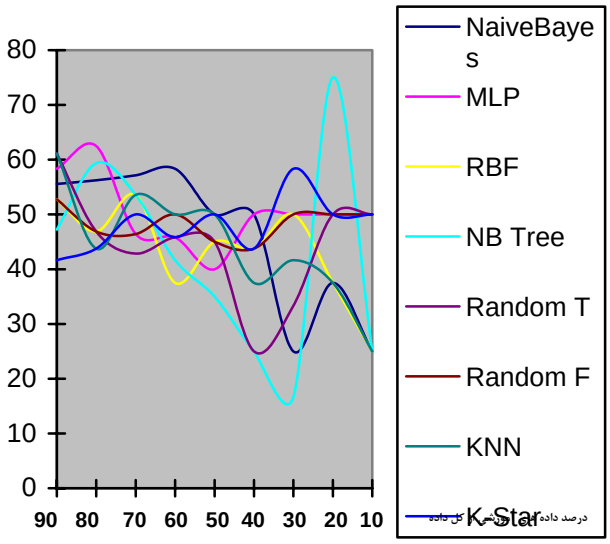
ما ابتدا حدود ۴۰ عکس را به صورت تصادفی انتخاب کردیم و در این ۴۰ عکس درصدی از داده ها را برای داده های آموزشی و درصد باقیمانده را برای داده های تست در نظر گرفتیم . چند نمونه از این داده ها را می توانید در شکل ۱ مشاهده نمائید .

شکل ۱ – نمونه ای از گوجه فرنگی های استفاده شده در مقاله

ما برای دسته بندی این داده ها از دسته بندی کننده های زیر استفاده کردیم :

۱. NaiveBayes
۲. Multi Layer Perceptron
۳. RBF Network
۴. Neural Binary Tree
۵. Random Tree
۶. Random Forest
۷. Instance Base K-Nearest Neighbor
۸. K-Star (K*)

نتایج بدست آمده از این روشها زمانی که داده های تست و آموزشی را با یک درصد مشخصی از کل داده ها در نظر بگیریم را می توانید در نمودار شماره ۱ و جدول شماره ۱ و جدول شماره ۲ ببینید . این نمودار بر پایه درصد دسته بندی درست برای هر روش بر حسب در صد داده های آموزشی از کل داده هاست .



نمودار شماره ۱ - نمودار درصد دسته بندی درست برای هر روش بر حسب درصد داده های آموزشی از کل داده ها

	Random Tree	Random Forest	KNN	KStar
90%	61.11	52.77	61.11	41.66
80%	46.87	46.87	43.75	43.75
70%	42.85	46.42	53.57	50
60%	45.83	50	50	45.83
50%	45	45	50	50
40%	25	43.75	37.5	43.75
30%	33.33	50	41.66	58.33
20%	50	50	37.5	50
10%	50	50	25	50

جدول شماره ۱ - درصد دسته بندی درست برای هر روش بر حسب درصد داده های آموزشی از کل داده ها

	Naive Bayes	MLP	RBF	NB Tree
90%	55.55	58.33	52.77	47.22
80%	56.25	62.5	46.87	59.37
70%	57.14	46.42	53.57	53.57
60%	58.33	45.83	37.5	41.66
50%	50	40	45	35
40%	50	50	43.75	25
30%	25	50	50	16.66
20%	37.5	50	37.5	75
10%	25	50	25	25

جدول شماره ۲ - درصد دسته بندی درست برای هر روش بر حسب درصد داده های آموزشی از کل داده ها

با دقت در نتایج بدست آمده می توان دریافت که اولا نتایج بدست آمده از نظم خاصی پیروی نمی کنند ثانیا درصد درستی نتایج نیز چندان مناسب نیستند .

برای فهم دلایل این موضوع ما یک بررسی اساسی را بر روی خصوصیات استفاده شده انجام دادیم. نتایج بدست آمده نشان می داد که برخی از خصوصیات اصلا تاثیری در نتایج بدست آمده ندارند بدین معنی که اگر این خصوصیات را حذف کنیم در نتایج دسته بندی تاثیری نخواهد گذاشت.

این خصوصیات عبارتند از :

۱. میانگین زردی و قرمزی و سبزی
۲. همگنی
۳. گردی
۴. مساحت

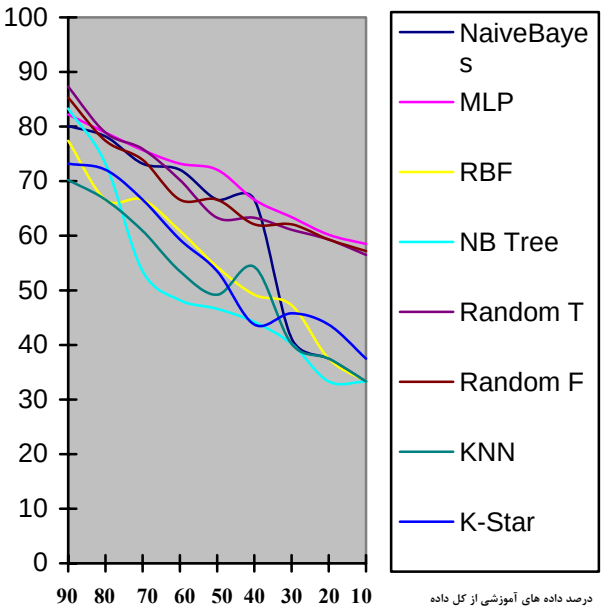
به طور کلی می توان نشان داد که این داده ها و این ترکیب خصوصیات برای دسته بندی چندان مناسب نیستند .

نکات منفی که می توان در مورد این داده ها نام برد عبارتند از :

۱. تعداد کم داده ها
۲. استخراج خصوصیات غیر ضروری
۳. استخراج نکردن خصوصیات ضروری از قبیل

به همین منظور ما حدود ۵۰ داده دیگر جمع آوری کردیم و همچنین با حذف خصوصیات غیر ضروری ذکر شده در بالا , خصوصیات مساحت خرابی سطح گوجه فرنگی و درصد مساحت خرابی به مساحت کل گوجه فرنگی را نیز در دسته بندی در نظر گرفتیم.

نتایج به دست آمده بهبود خوبی را نشان می داد که در جدول شماره ۳ و جدول شماره ۴ و نمودار شماره ۲ به خوبی قابل مشاهده است .



نمودار شماره ۲- نمودار درصد دسته بندی درست برای هر روش بر حسب درصد داده های آموزشی از کل داده ها

	Random Tree	Random Forest	KNN	KStar
90%	87.3	85.3	70.2	73.2
80%	78.9	77.4	66.6	72.1
70%	75.9	73.9	60.9	66.6
60%	70.2	66.6	53.5	59.3
50%	63.3	66.6	49.2	53.5
40%	63.3	62.1	54.3	43.75
30%	61.1	62.1	40.2	45.8
20%	59.3	59.3	37.5	43.7
10%	56.5	57.2	33.3	37.5

جدول شماره ۳ - درصد دسته بندی درست برای هر روش بر حسب درصد داده های آموزشی از کل داده ها

	Naive Bayes	MLP	RBF	NB Tree
90%	80.1	82.2	77.4	83.3
80%	78.2	78.9	66.6	73.2
70%	73.2	75.7	66.6	53.5
60%	72.1	73.2	60.9	48.2
50%	66.6	72.1	54.3	46.6
40%	66.6	66.6	49.2	44.2
30%	41.2	63.4	47.3	40.2
20%	37.5	60.2	37.5	33.3
10%	33.3	58.5	33.3	33.3

جدول شماره ۴ - درصد دسته بندی درست برای هر روش بر حسب درصد داده های آموزشی از کل داده ها

۳- نتیجه گیری

با توجه به نکات مختلف ذکر شده در این مقاله به طور کلی می توان نشان داد که برای دسته بندی داده های مربوط به گوجه فرنگی خصوصیات زیر تاثیر بیشتری دارند :

۱. سبز بودن
۲. قرمز بودن
۳. زرد بودن
۴. گردی
۵. مساحت خرابی سطح گوجه فرنگی
۶. درصد مساحت خرابی به مساحت کل گوجه فرنگی

همچنین با نگاهی به نتایج به دست آمده با استفاده از دسته بندی کننده های مختلف می توان نشان داد که MLP و Random Tree نسبت به دیگر دسته بندی کننده ها نتایج بهتری را نشان می دهند . همچنین می توان نشان داد که افزایش تعداد داده ها و در نتیجه افزایش تعداد داده های تست باعث افزایش کارایی دسته بندی کننده ها خواهد شد.

سپاسگزاری

در پایان جا دارد تـا از زحمـات اسـتاد عزیـزم جنـاب آقـای مهندس اکبری تشکر و قدردانی کنم .

مراجع

[1] Ruming Fang, Jiangrong Cai and Li Xu., *Computer image processing technique and applications in agricultural engineers,* Tsinghua university publications (1 999).

[2] N. Sarkar, R.R. Wolfe, *Feature extraction techniques for sorting tomatoes by computer vision,*TRANSACTIONS of the ASAE 28(3) 970-974 (1995).

[3] Cao, Masatem Nagata et al., *Studies on Grade Judgment of Tomato -Qualily Feature Extraction Using Color Image Processing,* Bulletin of the Faculty of Agriculture Miyazaki University 45(1-2)(127-133)(1998)

[4] Masatem Nagata, Pepito Menguito BATO, Masafui Mitarai, Qixin Cao, Taka0 Kitahara, *Study on Sorting System for Strawbeny Using Machine fision,* Journal of JSAM 62(1) (2000).

[5] *S.* Laykin, Y. Edan, V. Alchanatiset and R.Regev et al., *Development of a Quality Sorting Machine Using Machine fision and Impact,* ASAE Paper No.993 144, Toronto, Ontario, Canada July 18-2 1 (2002).

[6] YTao, P.H.Heinemann, Z.Varghese, C.T.Momw, H. J. Sommer, *Machine fision for Color Inspection of Potatoes and Apples,* TRANSACTIONS of the ASAE Paer No.993144 , Toronto, Ontario, Canada July 18-2 1 (2001).

[7] Gerhard Jahns1, Henrik Møller Nielsen2 and Wolfgang Paul, MEASURING IMAGE ANALYSIS ATTRIBUTES AND MODELLING FUZZY CONSUMER ASPECTS FOR TOMATO QUALITY GRADING , *Institute of Biosystems Engineering, Bundesallee 50, D-38116 Braunschweig - FAL, Germany(2004).*